DOCUMENT RESUME

ED 091 433 TM 003 647

AUTHOR Evans, John W.
TITLE Evaluating Education Programs: Are We Getting Anywhere?
PUB DATE [Apr 74]
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Educational Assessment; Educational Experiments; Educational Planning; Educational Policy; *Educational Programs; *Evaluation; Evaluation Needs; History; Political Issues; Program Effectiveness

ABSTRACT

This paper asks whether all the current attention being given to educational evaluation and all the activity going on indicates real progress in the output of evaluation and its use in the policy process. The paper reviews the brief history of educational evaluation and gives a qualified "yes" as an answer to the question, noting: significant progress in the funds and people being devoted to evaluation; improvement in the organizational location of the evaluation function in Federal agencies; increased use of more sophisticated evaluation methodology; the beginnings of the use of experimentation as a developmental precursor to the launching of national service programs; and the completion of a number of large-scale educational evaluations with major policy implications. The paper concludes by noting that despite real progress, serious administrative, methodological, and political problems threaten the continued expansion of evaluation studies and their use as a major factor in policy development and program administration. (Author)

Session # 1912

ED 091433

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

EVALUATING EDUCATION PROGRAMS--ARE WE GETTING ANYWHERE?

John W. Evans
Assistant Commissioner for
Planning, Budgeting, and Evaluation
U.S. Office of Education

TM 003 647

Invitational address before the American Educational Research Association
Chicago, April 18, 1974

ABSTRACT

This paper asks whether all the current attention being given to educational evaluation and all the activity going on indicates real progress in the output of evaluation and its use in the policy process. The paper reviews the brief history of educational evaluation and gives a qualified "yes" as an answer to the question, noting: sigrificant progress in the funds and people being devoted to evaluation; improvement in the organizational location of the evaluation function in Federal agencies; increased use of more sophisticated evaluation methodology; the beginnings of the use of experimentation as a developmental precursor to the launching of national service programs; and the completion of a number of large-scale educational evaluations with major policy implications. The paper concludes by noting that despite real progress, serious administrative, methodological, and political problems threaten the continued expansion of evaluation studies and their use as a major factor in policy development and program administration.

Evaluating Education Programs--Are We Getting Anywhere? [1]

John W. Evans
Assistant Commissioner for
Planning, Budgeting, and Evaluation
U.S. Office of Education

Now seems to be a time when basic reassessments are in order; so it is only appropriate for those of us concerned with educational evaluation to take stock of our own endeavors and try to answer the question I have posed: Are we really getting anywhere in our efforts to assess the effectiveness of educational programs, or is all the current talk and frenetic activity a case of much ado about nothing? The answer to this question is by no means obvious, even though it is the kind of rhetorical question that papers like this always ask, and answer with rosy if vague and overqualified bromides. Before you lean too far forward with anticipation, let me assure you that I'm going to hedge and qualify too, but my basic answer to the question is: "Yes, we are getting somewhere." I believe important progress has been made in recent years in educational evaluation in a variety of ways which I intend to specify, but the educational evaluation scene is not an untroubled one. Far from it. There are serious new problems that threaten the efforts of those of us who want to see the progress that has been made in educational evaluation continue, and I intend to talk about those also.

I. A Brief Look at History

It is appropriate that we begin with some sense of history, some understanding of where education evaluation has come and where it is today. We can begin that historical review with some very simple and chastening assertions.

---

1. Invitational address before the American Educational Research Association Chicago, April 18, 1974.

First, the history of significant evaluations of educational programs is brief and thin. In this respect educational evaluation is of a piece with the evaluation of social action programs generally. If we look back over the history of Federal efforts in the social program area, even back into the period of the New Deal and on up through the Great Society programs of the sixties, we are forced to acknowledge that virtually all of the original decisions by the Congress and the Executive Branch of the Federal government to initiate programs in the areas of education, manpower, and poverty, and the later decisions to continue, expand, or terminate these programs, were all taken with scarcely any empirical knowledge about the size, character, and location of the problems, or the likely effectiveness of proposed programs to remedy them. Once instituted, such programs were only rarely subjected to rigorous objective evaluation.

Second, the failure to evaluate education programs is a shortcoming not limited to the Federal Government. States and localities, which supply 95% of the funds for public education, have done virtually nothing to evaluate the effectiveness of their school systems and educational approaches.

Third, with a few notable exceptions, academic social scientists, traditionally accustomed to the research style of the individual scholarly grant and largely preoccupied with disciplinary issues and basic research, have made almost no contribution to actual evaluations of ongoing educational programs.

Such is the history of our efforts to formally evaluate most of our national education and other domestic programs. Yet, in less than ten years we have gone from a dearth of evaluation activity to a situation where

evaluation is now all the rage. Even though the amount of cocktail party conversation and the number of professional association meetings currently being devoted to evaluation are exaggerated indicators of the amount of useful, policy relevant evaluation which is now going on, there is no doubt that the change has been real and substantial.

What has accounted for this relatively sudden upsurge in attention to evaluation, and does it amount to real progress--progress in conducting sound evaluations and making them part of the policy process? There is no simple answer to this question, but from my vantage point at the Federal level there seem to be several important factors which have accounted for the increased concern with evaluation.

First of all, there is the long-run, cumulative effect of the presence and force of social science in our society. Social scientists have long been chiding administrators, policymakers, and Congressmen to rely less on subjective and political reasons for making decisions, allocating resources, and developing programs, and instead to make more use of the research methods and findings of social science. These entreaties have often turned out to be more rhetorical than substantive once the challenge was taken up. But they have not gone without effect on the policymakers, who have been made to feel increasingly guilty about not formulating policy and making decisions in a more rational way.

Second, there has been a gradual transformation in the intellectual make-up of the kind of people who have found themselves in both appointed and elected offices. While it remains true that raw, unreasoned, political interest still is the main factor in many decisions made by both the Executive Branch and

the Congress it is also true that the last 10 to 20 years have seen a significant increase in the number of people in such positions who want to attack a problem by asking what the real dimensions of it are, and how effective the available methods of treating it are--that is, people who want to try to rationalize the policy process.

Certainly another important factor in accounting for the upsurge of concern with evaluation at the Federal level was the implementation of the Program Planning and Budgeting System, or PPBS as it is usually called. As most of you know, this approach to analyzing and making decisions about program budget levels is a radical departure from the traditional incremental approach. It turns the focus away from the standard administrative budget categories toward the objectives, methods, and outcomes of programs. Such a shift automatically brings the need for data on program effectiveness to the fore.

Finally, there is the accountability movement itself in education. Obviously, it is hard to know whether the accountability movement is a cause, an effect, or merely an indicator of the increased interest in evaluation. In any case, once present, it has become a force in its own right.

But what have all these changes in analytical approach and expressed concern amounted to beyond heating up the atmosphere and expanding the rhetoric? What have they resulted in that allows one to conclude that some actual progress in educational evaluation is being made? To answer that question, let us look at three aspects of the evaluation process: first, the in-puts and resources; second, the methodology; and third, actual evaluation studies and their results.

II Some Indicators of Progress

A. Resources for Evaluation and the Avenues of Impact

There have been major increases in the wherewithall required for evaluations to get done, and important improvements in the organizational location of the evaluation function in the Government's decision-making apparatus. As the larger social changes I noted earlier have heightened concern with evaluation generally, the Congress has tired of listening to requests for increased appropriations based wholly on anecdotes and testimonials, and has increasingly demanded that Executive Branch Agencies produce some hard data on the effectiveness of their programs. For its part, the Congress has substantially increased funds and personnel to the domestic agencies for evaluation. In 1965, the Departments of Labor and HEW had available less than $5.0 million for program evaluation. By 1974 this figure had increased more than tenfold to more than $50.0 million.

There have been equally important changes in the organizational location of the evaluation function. As those of us who have worked in program agencies can testify, one of the indispensable prerequisites if evaluation is to impact on decisions is that it must be an integral part of top management's decision-making structure. Yet, as many of us also know, evaluation, like research, has often been buried in the bowels of program agencies and not only has gotten the leftovers in fiscal and personnel resources but has had little opportunity to make a meaningful input into the policy process. Even this is now changing. Most of the major Federal agencies now have an Assistant Secretary for Planning and Evaluation or its equivalent, and

the shoe is now on the other foot. Instead of having to plead for more money and people and the chance to participate in the decision-making process, evaluators are now under pressure to produce and to justify their claims of utility and relevance. As many of us have found out, demanding our place in the sun is a lot easier than justifying it.

Perhaps the most important thing to come out of all these resource increases and organizational changes is that not only is it now possible to do evaluations and have them taken seriously, but the basic dialogue of management has begun to change from considerations of how big a program's budget should be and the constituency pressures for its continuation, to considerations of objective evidence of performance and indicators of program effectiveness.

B. The Use of More Sophisticated Methodology in Evaluation

In addition to these increases in resources for evaluation and improvements in the opportunities for its use in the policy process, there have been some important methodological advances that should not be taken lightly. I want to touch briefly on just two. The first is the appearance of efforts in large scale national evaluations to use the classic model of experimental design with randomized treatment and control groups.

Since virtually all education programs have their committed advocates and strong detractors, and are in this sense inherently controversial, it is inevitable that evaluations of them will also be controversial. Any evaluation which finds a program successful will be attacked by the program's detractors, and any evaluation which finds the program unsuccessful will be

denounced by its advocates. As Peter Rossi has put it, "No good evaluation goes unpunished."[2] In such cases, the controversy will not be waged directly over what is truly at stake, namely, the disagreeableness of the findings, but instead will take the form of an attack on the validity of the evaluation's methodology. Acrimonious debates will rage through the pages of the press and the professional journals over sample size inadequacies, non-representativeness, culture-biased measurement instruments, failures to meet the assumptions of parametric statistical models, and the like. Since no evaluation can ever be flawless, especially those carried out in the real world of classrooms and communities, evaluators will never escape these post evaluation debates--nor, indeed, should they. But to strengthen the validity of evaluation findings and the justification for using formal empirical evaluations as a basis for policy decisions, it is important that they be as methodologically strong as possible.

The major weakness in most evaluation designs relates to the use of control groups. The feature of education evaluations that has proven most vulnerable to both well-motivated and not so well-motivated attack is the comparability between treatment and control groups. Once evaluations move beyond the primitive efforts to conduct site visits or simply collect data on a before-after basis, the evaluator and the design he employs must confront the fundamental problem of providing some estimate of what would have happened in the absence of the program he is evaluating. There are of course a variety of ways to

---

2. Peter Rossi, "Testing for Success and Failure in Social Action," in Rossi and Williams, Evaluating Social Programs, New York: Seminar Press, 1972, p. 32.

deal with this problem, which include comparison with national norms, comparison with previous years' scores, the use of a matched comparison group, etc. But the effects of education programs are seldom dramatic, and the small differences they are likely to make can easily be either overestimated or missed entirely by comparing the treatment group with a non-comparable control group. The only truly satisfactory way of dealing with this problem, as we all know, is through the classic experimental design model, with randomly assigned treatment and control groups. Don Campbell, both in his paper on "Reforms as Experiments"[3] and elsewhere, has written eloquently and extensively on this issue, urging that educational and other kinds of programs be structured at the outset to allow this kind of evaluation design. He and I have debated the question of how acceptable evaluations are which fall short of this standard.[4] My view is that there will be many instances where this obviously preferable design is not feasible, and that rather than throw up our hands and withdraw from the arena because we cannot have random assignment, we must carry out whatever kind of evaluation is feasible and useful within the time constraints of the policy process and make the best use of it we can. It is my experience that even fairly primitive designs are likely to provide better data for decision-making than the subjective impressions and partisan arguments normally used.

---

3. Donald T. Campbell, "Reforms as Experiments," American Psychologist. Vol. 24, No. 4, April 1969, pp. 409-429.

4. Donald T. Campbell and Albert Erlebacher, "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful," in J. Hellmuth, Ed, Compensatory Education: A National Debate, Vol. 3, Disadvantaged Child, New York: Brunner/Mazel, 1970

   John W. Evans and Jeffry Schiller, "How Preoccupation with Possible Repression Artifacts Can Lead to a Faulty Strategy for the Evaluation of Social Action Programs: A Reply to Campbell and Erlebacher," J. Hellmuth, op. cit.

In any case, my point here is simply to say that while we must be willing to do what we can to improve the decision-making process, I nevertheless agree with Campbell that we should certainly strive to use the superior method of randomized experimental design whenever we can. And I would report that the Evaluation Unit at the Office of Education has been able to mount two such randomized design evaluations of the Emergency School Assistance Program in the last two years. One is completed and the other is still in process. In a forthcoming paper in which he analyzes efforts to introduce and assess innovations across a braod variety of fields from medicine to education, Frederick Mosteller[5] notes that this is apparently the first time in education in which a major evaluation was carried out using this type of design.

The other methodological advance I want to say a word about is the emergence of experimentation as a precursor to the full fledged introduction of major new programs. This could be one of the most important developments in our time if it takes hold and is actually used--and those are two big ifs.

In order to understand what social or educational experimentation entails and how it may be of great value in developing educational policies and programs, it may be helpful to look at what has happened in the field of compensatory education. In the early sixties the country belatedly recognized the existence of the disadvantaged child. It was acknowledged that about a fifth of our children arrive at the first grade with educational deficits which are measurable even at that time; that as they progress through school the achievement gap

---

5. Frederick Mosteller, Social Experimentation (forthcoming)

between them and their middle class peers widens; that a great many do not learn to read, write, or calculate adequately; that they drop out of school in large numbers; and that as a consequence of these deficits in basic skills and credentials they are unable to pursue postsecondary education or form either a lasting or satisfying attachment to the labor market. What is worse, these educational disadvantages are passed on culturally to their children, and a significant portion of the society's population is caught up in a cycle of intergenerational poverty.

Once these problems were recognized, the sense of political urgency was irresistible, and the country rushed to pass major legislation and initiate, among other things, a number of major early childhood compensatory education programs. In this atmosphere of rushing to solve the problem, only the faintest of voices were heard asking the unaskable questions: Did we really know what we were doing? Did we really have effective program models and techniques which could remediate the educational deficits of these children?

The evaluations which have since been carried out on these national programs, principally Head Start and Title I, indicate that we did not know what we were doing.[6] While our ideology was laudable and our motivations pure, our programmatic know-how was skimpy.

The problem is that once large national programs are put into place, the political force of their authorship and the pressure from their constituencies

6. See, for example: Westinghouse Learning Corporation, The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development, Vols. I and II, June 12, 1969; and M. Wargo, et al, ESEA Title I: A Reanalysis and Synthesis of Evaluation Data from 1965-1970, American Institutes of Research, Palo Alto, California, March 1972.

for continued funding make it almost impossible to even acknowledge publicly that they may not be effective, much less attempt to alter them in some fundamental way to make them so. This is where a strategy of experimentation should be of great value. After there is broad recognition of a major educational or social problem, if we can force ourselves to recognize that we may not know how to solve it, and if instead of going directly to a massive national program, we will initiate a controlled experiment in which we develop and test the relative effectiveness of alternative programmatic techniques, we can reap a number of benefits. First, if the results of that experiment show that we have not yet achieved effective program models and techniques, it is politically possible to admit this and go back to the drawing board to develop them. Moreover, we can take considerable satisfaction in the knowledge that we are not committed to the continued expenditure of large resources on efforts we know to be ineffective. On the other hand, if the experiment is successful, we can go forward with a large national service program reasonably confident that the massive resources we will be devoting to the problem will have a good chance of actually making a dent in it.

The logic of using experimentation as a developmental precursor to national program implementation may seem compelling enough, but the rush to adopt this strategy has not exactly been a stampede. In fact, my colleague Michael Timpane has made a series of observations on the potential of social experimentation which I have jokingly referred to as Timpane's law.[7]

7. Michael Timpane, "Educational Experimentation in National Social Policy," Harvard Educational Review, Vol. 40, No. 4 pp. 547-566, Nov. 1970.

His conclusion is that if there is enough interest in some problem to support a major social experiment, then the interest will be so great that no one will be willing to wait for the conclusion of the experiment before passing legislation to implement a national program. On the other hand, if there is not broad concern over the problem, then there won't be enough interest in Congress to support the funding of an experiment on it. Either way there is no experiment.

I wish I could say that this "law" was no more than humorous by-play. Even though there is truth as well as humor in Timpane's law, some significant efforts at experimentation in education are nevertheless occurring, and we should not overlook their importance both as early prototypes of what could become a fundamentally new way of approaching the development and initiation of education programs, and for the contribution they have made to the particular educational issues they address. I would cite two such efforts. The first is the Follow Through program and the second is the OEO experiment on Performance Contracting.

In the case of Follow Through, this program was originally intended to be a follow-up service program intended to reinforce whatever gains were made in Head Start; but by the time its first appropriation had passed through the various budget cutting phases, the initial request of $120 million in 1968 had been reduced to $15 million. Realizing that it made no sense to mount a service program which could address only one percent of the target population, the program staff shifted the focus of the program away from service delivery to the development and evaluation of alternative compensa-

tory education models. This effort at educational experimentation has certainly not been an exemplary one. It has been plagued by staff shortages, administrative difficulties, and continued unclarity over what Follow Through's true mission is and how it should be carried out. Nonetheless, despite these problems, some evaluation findings are now beginning to emerge which are precisely the kinds of outcomes we would expect from a planned variation experiment. Some of the program models are showing the ability to produce cognitive and affective gains that are larger than those we have seen in most compensatory education programs. Other models are producing gains that are just about what one would expect from the normal school experience, while still others are apparently so ineffective that the children in the control group are educationally better off than those in the model programs. If these findings hold up in the subsequent waves of the longitudinal evaluation, we should have a much better basis on which to proceed programmatically in the area of early childhood compensatory education.

The OE experiment on Performance Contracting grew out of the kind of situation which should call for an experiment. You will recall that about four years ago a number of educational technology firms were promoting the ability of their techniques to produce large gains in reading and math among disadvantaged children. Interest in performance contracts began to sweep through public school systems with large populations of educationally disadvantaged children. The attraction to performance contracting was based on a number of factors. It was at this time that the disillusionment about

public education which flowed from the Coleman and later analyses was approaching its peak. The siren of performance contracting was especially seductive at this time because it said: "Not only is it possible to remediate the deficits of disadvantaged children, but we have the techniques to do it, we are ready to come into your schools and implement it, it is no more expensive than your present per pupil expenditure, and we will sign a binding contract with you which says that if we don't produce significant, independently measured gains in reading and math, you don't have to pay us." Small wonder that these blandishments triggered a rush to the performance contractors' door.

But there were also strident critics of performance contracting, mainly the teachers' unions, who argued that performance contracting was an illusory panacea and that it would dehumanize the learning process. Depending on who won the argument--that is, who shouted the loudest--it seemed that performance contracting was destined to be either prematurely buried or unjustifiably expanded into a national movement.

Noting that these unfounded claims and counter charges were precisely the circumstances which call for an experiment, the evaluation staff at OEO designed and carried out such an experiment, underwriting and independently evaluating seven different performance contracting firms. The results of the evaluation, as most of you know, showed that none of the performance contract models was able to produce reading and math gains that were significantly better than the results achieved through the regular public school methods. It is hard to predict what the outcome of the debate would have been had the experiment not been done.

C. Studies and Results

Given that educational evaluation has shown progress in the funding support it has attracted, in the sheer amount of evaluation activity that is going on, in the organizational position that evaluators hold in Government agencies, in the demand for evaluation results by the Congress, and in the use of more sophisticated designs in the conduct of evaluations, what has actually been done by way of major evaluation studies that have important policy implications? It is the completion of such actual studies, after all, that is the outcome measure for evaluating evaluation.

First of all, there is the Coleman Report itself which, while it does not evaluate a specific educational program, nevertheless is fundamentally an evaluative analysis assessing the effects of what had traditionally been regarded as some of the most important independent variables in the educational process. Notwithstanding the continuing debates over the methodological shortcomings of the Coleman Report, few would deny that it is a landmark study which has caused educational theorists to reassess their fundamental beliefs and strategies and legislators to reexamine their unquestioned faith in educational programs and appropriations.

Second, the OEO evaluation of Head Start, usually referred to as the Westinghouse Report, which was also the subject of intensive methodological scrutiny and debate, is one of a number of studies of early childhood compensatory education program which shook us out of our complacent belief

that popular, well motivated programs for poor kids are necessarily effective in remediating their educational deficits.

Third, an evaluation of the Emergency School Assistance Program found that this moderately funded and locally generated collection of projects was able to significantly increase the achievement levels of black male teenagers, and thus, by indicating that compensatory education in the public schools is possible, was a welcome contradiction to the largely negative findings of so many of the earlier studies.

Fourth, and in the same vein, an early evaluation of the Upward Bound program found that this program was effective in persuading low income high school youngsters to attend college, in keeping them there, and in graduating them at a rate which made the program cost-beneficial.

I have already mentioned the major evaluations of the Follow Through Program and Performance Contracting. I don't wish to extend this list indefinitely, mainly because I couldn't, even if I wanted to. But I do want to make the point that if we ask whether all the hoopla of educational evaluation has amounted to anything more than increases in funds, data gathering, and professional meetings, the answer is yes. There is far less on the production ledger of educational evaluation than there should be, but indications of important progress are by no means lacking.

III. A Look at the Future: Prospects and Problems

This recitation of progress makes things sound a lot better than they are. To be sure, the progress is real; but in the last few years a number of

new problems have arisen which the practitioners of educational evaluation will have to solve if they wish to see the use of evaluation in the policy process progress beyond its present promising but inchoate state. I call them new problems because it is important to distinguish them from the kind of problems we would have listed ten years ago. The review I have made so far should make it clear that educational evaluators can no longer complain that they do not have enough money or people or that they are not taken seriously by policymakers and legislators. Moreover, I do not agree with those who argue that methodological inadequacies of one sort or another present a major obstacle to the full flowering of educational evaluation as a policy instrument. It is not uncommon for social scientists to display handwringing despair over their primitive methods and insensitive measuring instruments, and to plead that an Einsteinian breakthrough in the social sciences is needed to put things right. My own view is that we have a long way to go in making full use of the techniques we have before we are in a position to complain about inadequate methods.

The newer problems which education evaluators face are of a different order, and I will try to indicate what I think some of them are:

1. As educational research and evaluation have proliferated, the people and institutions who are the objects of these studies have come under an increasing data collection burden--and are more and more expressing their resistence to it. It is no longer possible for evaluators to assemble

a battery of interview schedules and questionnaires and invade the schools. Extensive prior clearance and review are now required almost everywhere, and outright refusal to participate in studies is not uncommon. The research and evaluation community is going to have to work out some collective way of dealing with this problem, for it is a real one. By the time school systems total up all the data collection requirements which come from Federal, State, local, and private requirements, the burden often is a crushing one.

2. Evaluation studies which involve collecting data on adults are encountering increasing resistance at the interviewee level, particularly among minorities and the poor where it is now not uncommon for respondents to insist that they be paid for their time.

3. The increased sensitivity to evaluation studies--both what they seek to find out and the amount of data they propose to collect--is resulting in a strangling growth of reviews, clearances, and advisory bodies. The problems which these multiple involvements and clearances pose for the evaluator are so great that it threatens to prevent many evaluations from being carried out at all.

4. As protests over evaluations arise, ostensibly because of objections to the type and amount of data to be collected, there is

likely to be an increased politicization of these protests and their use as weapons in broader disputes between State and local officials, between school administrators and unions, or between local and Federal levels of government.

5. As evaluation activity and policymakers' interest in it have grown, there has also been an increased awareness at the program level that it is necessary to start taking evaluations seriously. This has had the unfortunate effect on some program officers and school administrators of increasing their unwillingness to participate in evaluation studies because of their fear of what will happen to their programs if the evaluation produces negative findings.

6. Evaluators are increasingly subject to unrealistic expectations on the part of policymakers and legislators with respect to both the speed with which evaluations should be mounted and completed, and the simplicity of the answers which are desired. Having whetted the appetite of decisionmakers, a demand has been created and it is an increasingly insistent one. Policymakers are beginning to display an irritated impatience with the elaborate trappings of careful design, longitudinal studies, and complex multivariate findings. They want to know whether or not a program is any good and they want to know it yesterday. As unrealistic as these expectations are, evaluators themselves

probably must bear some of the blame for them. In their early zeal to have the virtues of evaluation recognized and used by policymakers, evaluators were almost certainly guilty of overpromising.

This problem has already gotten beyond the stage of irritated impatience. Last year, the Congress cut the Office of Education's evaluation budget in half and made large reductions in its statistics budget and in NIE's research funds.

7. We are certain to see a lot more public debate of the kind I spoke of earlier over the validity of evaluation methodology and its results; and an increasingly important and time consuming task for evaluators will be defending the evaluations they carry out and their suitability as a basis for policy decisions. An unfortunate by-product of such debates is the impression created among both policymakers and the public that the mere fact such a debate is occurring means the evaluation must ipso facto be faulty and therefore should be put aside. It is ironic that after a large scale formal evaluation has been put aside because of technical questions raised about its methodology, policymakers and program officials then return to the old and familiar methods of making the decision or formulating the policy--methods which are totally partisan and subjective in nature.

The seriousness of these problems should not be underestimated merely because so many of them are technical and procedural in character. Perhaps evaluators

can take some solace, however, in the realization that these are the problems of impact and success rather than the problems of neglect and disregard.

The fact that evaluators are now facing such problems is an indication of how far evaluation has come in the last decade. Educational evaluation has gone from not being taken seriously to being expected to produce. It has gone from a condition of no funds, people, or influence to one of being held accountable for producing valid and useful studies. It has gone from not having enough money to do evaluations at all to the technical and political problems of carrying them out. Some evaluators who have struggled so hard to bring about these changes are now wistfully wondering whether they wouldn't just as soon have their old problems back. As Oscar Wilde observed, there are two particularly dissatisfying things in life: the first is not getting what you want; the second is getting it.

Finally, in sum, while I do not agree with the cynical view which holds that educational evaluation is largely a waste of time because its methods are too weak, because it will be forever undersupported, or because important policies and decisions will be made in spite of evaluation findings, and while I believe that important progress has been made in educational evaluation during the past decade--in increased support and opportunities for influence, and in important substantive results--I nevertheless believe that educational evaluation now faces a new array of problems that are possibly more serious than the basic ones of getting the necessary resources to do evaluations. These new problems are a strange mixture of logistics and politics, and they are in large

part ar outgrowth of the increasing pluralism of American society. If these problems are not dealt with, evaluation will not succeed in making more than an occasional or marginal impact on educational policies and programs. If these problems are solved, the general trend, which has only recently been established, can be continued; and the wider use of evaluation can make a major contribution to the setting of national educational policies, to the development of education programs, and to the allocation to scarce educational resources.